ƯỚC LƯỢNG

Ước lượng điểm

- Ví dụ: công ty A có hàng ngàn công nhân.
  - Thăm dò 100 công nhân của công ty nhận thấy thu nhập trung bình là 1,5 triệu đồng/tháng.
  - Sử dụng trung bình mẫu để ước lượng thu nhập trung bình của công nhân công ty A.
  - Ta nói thu nhập trung bình của công nhân công ty được ước lượng là 1,5 triệu đồng/tháng,
- Một ước lượng tham số của tổng thể được cho bởi **một con số** thì được gọi là ước lượng điểm của tham số tổng thể.

Ước lượng khoảng

- Nếu nói một độ dài nào đó là 5,28 m thì ta nói về ước lượng điểm.
- Nếu nói độ dài đó là 5,28 $\pm$ 0,03 m, i.e độ dài đó nằm trong khoảng 5,25 và 5,31 m thì ta đã đưa ra một ước lượng khoảng.
- Giá trị 0,03 được gọi là **sai số** của phép ước lượng.
- Sai số của phép ước lượng, ký hiệu ε, ***thể hiện*** độ chính xáccủa phép ước lượng đó.
- Sai số càng nhỏ, độ chính xác càng cao.

Các khái niệm và kí hiệu chung cho bài toán ước lượng

Xét ĐLNN X trong tổng thể.

Xét một mẫu ngẫu nhiên gồm n phần tử.

μ: giá trị trung bình của ĐLNN X trong tổng thể.

$$\boxed{\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

trung bình mẫu cụ thể

σ: độ lệch chuẩn của X trong tổng thể.

s: độ lệch chuẩn có hiệu chỉnh của mẫu cụ thể

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1}}$$

Ước lượng trung bình tổng thể μ

Ước lượng trung bình tổng thể μ

- **Mục tiêu:** ước lượng μ với độ tin cậy 1- α
  – Độ tin cậy 1- α là khả năng μ thuộc vào khoảng tin cậy mà ta đưa ra.
  – Dùng giá trị trung bình của mẫu $(\overline{x})$ để ước lượng trung bình của tổng thể μ
- Ta cần tìm khoảng tin cậy $[\overline{x}-\varepsilon;\overline{x}+\varepsilon]$ sao cho:

$$P(\overline{x}-\varepsilon \leq \mu \leq \overline{x}+\varepsilon)=1-\alpha$$

- Ví dụ: Khoảng tin cậy là [156; 173] với độ tin cậy là 95% nghĩa là có 95% khả năng μ thuộc khoảng [156; 173] và có 5% khả năng μ không thuộc khoảng [156; 173].

Ước lượng μ với độ tin cậy 1-α

1) TH: σ đã biết (X có PP chuẩn, hoặc X không có PP chuẩn nhưng n >=30)

$$Z = \frac{\overline{X} - \mu}{(\sigma / \sqrt{n})} \quad \sim \quad N(0,1)$$

Tìm $z_{\alpha/2}$ sao cho $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$,

nghĩa là tìm $z_{\alpha/2}$ (tra bảng A4) thỏa

$$P(Z \le z_{\alpha/2}) = 1 - (\alpha/2)$$

Tính độ sai số của ước lượng $\boxed{\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$$[\overline{X} - \varepsilon, \overline{X} + \varepsilon]$$

Ước lượng μ với độ tin cậy 1-α

2) TH: σ chưa biết

a) n ≥ 30 : thay σ bởi s (X có phân phối bất kì)

$$Z = \frac{\bar{X} - \mu}{(s/\sqrt{n})} \sim N(0,1)$$

Tìm $z_{\alpha/2}$ (tra bảng A4) thỏa

$$P(Z \le z_{\alpha/2}) = 1 - (\alpha/2)$$

Tính độ sai số của ước lượng $\boxed{\varepsilon = z_{\alpha/2} \frac{s}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$$[\bar{X} - \varepsilon, \bar{X} + \varepsilon]$$

Ước lượng μ với độ tin cậy 1-α

2) TH: σ chưa biết

b) n < 30, X có phân phối chuẩn :

Thay σ bởi độ lệch chuẩn mẫu có điều chỉnh s

$$T = \frac{\bar{X} - \mu}{(s/\sqrt{n})}$$ Phân phối Student bậc tự do n-1

Tìm $t_{\alpha/2}$ (tra bảng A5, dòng n-1) thỏa

$$P(T > t_{\alpha/2}) = \frac{\alpha}{2}$$

Tính độ sai số của ước lượng $\boxed{\varepsilon = t_{\alpha/2}\frac{s}{\sqrt{n}}}$

Khoảng tin cậy cho μ với độ tin cậy 1-α:

$$[\bar{x} - \varepsilon, \bar{x} + \varepsilon]$$

Ước lượng μ với độ tin cậy 1-α

TH: $n < 30$ và X không có phân phối chuẩn

Tham khảo

Section 10.3.3 Bootstrap confidence intervals, Probability and Statistics for Computer Scientists, Michael Baron, second edition.

Ví dụ: mẫu lớn

Độ đo các đường kính của một mẫu ngẫu nhiên gồm 100 vòng bi do một máy sản xuất trong một tuần có đường kính trung bình là 0,824 cm và độ lệch chuẩn mẫu có điều chỉnh là 0,042 cm.
Hãy tìm khoảng tin cậy đường kính trung bình của tất cả các vòng bi với độ tin cậy 96%.

Ví dụ: mẫu bé, biết σ

- Trọng lượng của những vỉ thuốc do một công ty dược sản xuất ***có phân phối chuẩn*** với ***độ lệch chuẩn 0.038mg***. Một mẫu ngẫu nhiên gồm 10 vỉ thuốc có ***trọng lượng trung bình là 4.87mg***. Với ***độ tin cậy 95%,*** hãy ước lượng trọng lượng trung bình các vỉ thuốc do công ty sản xuất.

Ví dụ: mẫu bé, không biết σ

Một mẫu gồm 10 độ đo đường kính của một quả cầu, có đường kính trung bình là 4,38 cm và độ lệch chuẩn mẫu có điều chỉnh là 0,06 cm. Biết rằng đường kính của một quả cầu có *phân phối chuẩn*, hãy tìm khoảng tin cậy của đường kính tất cả các quả cầu với độ tin cậy 95%.

Ví dụ 5.7: Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên phân bố theo quy luật chuẩn với độ lệch tiêu chuẩn 1 gram. Cần thử 25 sản phẩm loại này ta thu được kết quả:

Trọng lượng (gram)	18	19	20	21
Số SP tương ứng	3	5	15	2

Với độ tin cậy 95%

a. Hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.

b. Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì cần cân thử ít nhất bao nhiêu sản phẩm.

Ví dụ 5.9: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $N(\mu;\sigma^2)$. Gieo thử giống hạt này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173,174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170.

Hãy tìm khoảng tin cậy cho năng suất trung bình của loại hạt giống này với độ tin cậy $\beta = 95\%$.

Giải: Năng suất trung bình của hạt giống là tham số μ.

Từ các số liệu trên ta tính được: $\overline{x} = 171$; $s = 3,4254$. $\alpha = 0,05$; $\frac{\alpha}{2} = 0,025$.

Tra bảng phân bố Student với 15 bậc tự do ta tìm được $t_{\alpha/2}(n-1) = t_{0,025}(15) = 2,131$.

Độ chính xác $\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} = 2,131 \cdot \frac{3,4254}{\sqrt{16}} = 1,885$.

Vậy khoảng tin cậy cho năng suất trung bình của loại hạt giống này là μ thỏa mãn:

$$169,115 \le \mu \le 172,885.$$

Ước lượng tỉ lệ

Ước lượng tỉ lệ tổng thể

Kí hiệu:

p: tỷ lệ phần tử có tính chất $\mathscr{P}$ trong tổng thể,

f: tỷ lệ phần tử có tính chất $\mathscr{P}$ trong mẫu.

Lưu ý: bài toán ước lượng tỷ lệ chỉ xét trong trường hợp mẫu lớn, do đó ta dùng phân phối chuẩn.

Ước lượng tỷ lệ tổng thể p với độ tin cậy 1- α

Khoảng tin cậy: $[f-\varepsilon\ ;\ f+\varepsilon]$

Độ chính xác (sai số)

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}$$

Trong đó $z_{\alpha/2}$ thỏa

$P(Z \le z_{\alpha/2}) = 1-(\alpha/2)$, $Z \sim N(0,1)$

(tra bảng A4 để xác định $z_{\alpha/2}$)

Bài toán xác định cỡ mẫu

Bài toán: Xác định cỡ mẫu cần khảo sát để đạt được ước lượng cho tỷ lệ tổng thể với độ tin cậy và độ sai số cho trước.

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \le z_{\alpha/2}\frac{1}{2\sqrt{n}} \qquad \forall \quad f \in [0,1]$$

$$\Rightarrow \qquad n \ge \left(\frac{z_{\alpha/2}}{2\varepsilon}\right)^2$$

Ví dụ

Thăm dò ý kiến của 100 cử tri được chọn ngẫu nhiên tại một địa phương cho thấy có 80% trong số này ủng hộ ứng cử viên A.
Với độ tin cậy 98%, hãy ước lượng tỷ lệ của tất cả các cử tri tại địa phương này ủng hộ ứng cử viên A.

Nhận xét về khoảng tin cậy

- Tung một đồng xu 1000 lần, thấy có 536 lần xuất hiện mặt ngửa.
- Với độ tin cậy 95% ta tìm được khoảng tin cậy
$$50.5\% < p < 56.7\%$$
- Nhận xét: **có đến 95% khả năng**tỉ lệ tổng thể p nằm trong khoảng từ 50.5% đến 56.7%
- Do đó **nhiều khả năng**, đây là một đồng xu không công bằng.

Dùng khoảng tin cậy của p để ước lượng N

Để điều tra số cá trong hồ, cơ quan quản lý đánh bắt 300 con, làm dấu rồi thả xuống hồ. Lần hai bắt ngẫu nhiên 400 con thấy có 60 con được đánh dấu. Hãy xác định số cá trong hồ với độ tin cậy 95%.

- $p = \frac{M}{N} = \frac{300}{N}$ tỷ lệ cá ***được đánh dấu*** trong hồ.
- $f = \frac{m}{n} = \frac{60}{400} = 15\%$: tỷ lệ cá được đánh dấu trong mẫu ngẫu nhiên gồm $n = 400$con cá.
- Với độ tin cậy 95%, sử dụng phân phối chuẩn để ước lượng p, sau đó ước lượng N.

Bài tập tương tự

- Một nhà sản xuất que diêm tuyên bố rằng 80% những hộp diêm do họ sản xuất có không ít hơn 50 que diêm.
- Để kiểm chứng điều này một khác hàng chọn ngẫu nhiên 250 hộp diêm và thấy rằng chỉ có 183 hộp có từ 50 que diêm trở lên.

a) Tìm khoảng tin cậy của tỷ lệ những hộp đạt tiêu chuẩn mà NSX tuyên bố với độ tin cậy 98%.

b) Với số liệu trên, bạn nhận xét gì về lời tuyên bố của nhà sản xuất ?

Đánh giá sai số của bài toán ước lượng tỉ lệ

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \le \frac{z_{\alpha/2}}{2\sqrt{n}} = \varepsilon_{\max}$$

Vì lí do tài chính, một tờ báo quyết định chỉ khảo sát 2000 cử tri về dự định của họ trong cuộc bầu cử sắp tới. Với độ tin cậy 95%, sai số tối đa khi ước lượng tỷ lệ cử tri bầu cho ứng viên X là bao nhiêu ?

Bài tập tương tự

A researcher wishes to estimate, with a probability of 0.95, the proportion to within 3% of mosquitos which carry a virus. How large must the sample be?

Một nhà nghiên cứu muốn ước lượng tỷ lệ những con muỗi mang virus sốt huyết với ***sai số tối đa*** là 3% và độ tin cậy 95%. Hỏi cần khảo sát mẫu có kích thước **nhỏ nhất** là bao nhiêu?

Ôn tập 1

Khảo sát doanh số bán của một siêu thị trong một số ngày được chọn ngẫu nhiên, ta thu được số liệu như bảng kế bên:

1. Hãy ước lượng doanh số bán trung bình của siêu thị trong một ngày với độ tin cậy 95%.
2. Nếu muốn sai số ước lượng doanh số trung bình trong ngày có sai số không quá2 triệu đồng/ngày ở độ tin cậy 99% thì có cần thêm ngày quan sát không? Nếu có thì cần quan sát thêm ít nhất bao nhiêu ngày ?
3. Những ngày có doanh số bán trên 130 triệu đồng được gọi là ngày đắt hàng. Hãy ước lượng tỷ lệ những ngày đắt hàng của siêu thị với độ tin cậy 90%.
4. Hãy ước lượng doanh số trung bình của một ngày đắt hàng với độ tin cậy 96%. Cho biết doanh số bán hàng có phân phối chuẩn.

Doanh số (triệu đồng)	Số ngày
50 - 70	6
70 - 90	30
90 - 110	40
110 - 130	25
130 - 150	15
150 – 200	5

Ôn tập 2

Giả sử tại tp HCM, thăm dò 500 người dân về việc đập bỏ Dinh Thượng Thư, có 380 người không đồng ý.

1. Với độ tin cậy 95%, hãy ước lượng tỷ lệ người dân tp HCM không đồng ý việc đập bỏ Dinh Thượng Thư
2. Nếu muốn độ chính xác của phép ước lượng là 3% thì độ tin cậy là bao nhiêu?
3. Nếu muốn phép ước lượng có độ tin cậy là 99% và độ chính xác là 3% thì cần thăm dò thêm bao nhiêu người nữa?